ΚΕΦΑΛΑΙΟ 8
ΤΑΞΙΝΟΜΗΤΕΣ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Σκοπός

Σκοπός του παρόντος κεφαλαίου είναι να εισαγάγει τον αναγνώστη σε μία σχετικά νέα κατηγορία ταξινομητών, γνωστών ως νευρωνικά δίκτυα. Τα δίκτυα αυτά αποτελούν σήμερα ένα από τα βασικά εργαλεία σχεδιασμού ταξινομητών.

Προσδοκώμενα Αποτελέσματα

Όταν θα έχετε τελειώσει τη μελέτη του κεφαλαίου αυτού θα μπορείτε:

- Να επιλύσετε ένα πρόβλημα ταξινόμησης με τον Αλγόριθμο Perceptron, όταν οι κλάσεις είναι γραμμικά διαχωρίσιμες.
- Να σχεδιάζετε το κύκλωμα ενός νευρώνα ή την Αρχιτεκτονική Perceptron, που μπορεί να υλοποιήσει τον ομώνυμο αλγόριθμο.
- Να σχεδιάζετε Perceptrons δύο ή τριών στρωμάτων για την επίλυση προβλημάτων ταξινόμησης με δύο ή περισσότερες κλάσεις.
- Να περιγράψετε τη βασική μορφή του αλγορίθμου Οπισθοδρομικής Διάδοσης, που χρησιμοποιείται στην εκπαίδευση πολυστρωματικών Perceptrons,

Έννοιες-κλειδιά

- Αλγόριθμος Απότομης Κατάδυσης
- Αλγόριθμος Εκμάθηνσης
- Αλγόριθμος Οπισθοδρομικής Διάδοσης.
- Αλγόριθμος Perceptron
- Γραμμικά Διαχωρίσιμες κλάσεις
- Εκπαίδευση του Perceptron
- Κρυφό Στρώμα Νευρώνων
- Λογιστική Συνάρτηση.
- Μηχανή που Μαθαίνει.
- Νευρώνας
- Συνάρτηση Ενεργοποίησης ενός Perceptron
- Συνάρτηση Κόστους
- Σύναψη
- Ταξινομητής Perceptron

Εισαγωγικές Παρατηρήσεις

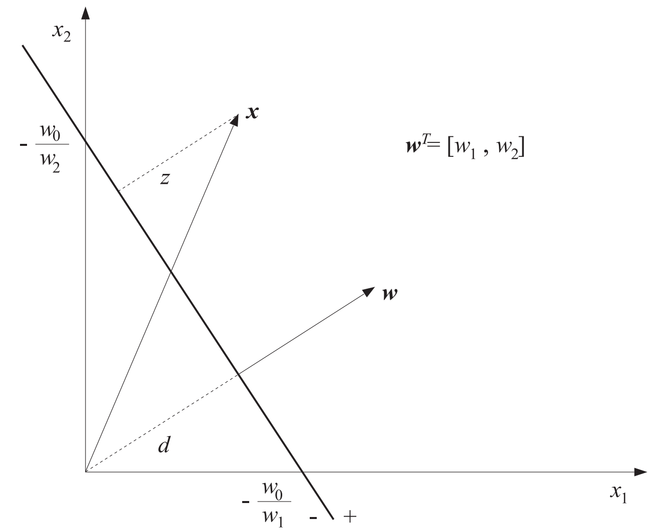
Όπως έχουμε ήδη αναφέρει στο προηγούμενο κεφάλαιο, ο σκοπός ενός ταξινομητή είναι να χωρίσει το χώρο σε περιοχές και να τις “ονοματίσει” σύμφωνα με τις υπάρχουσες κλάσεις. Ο βέλτιστος τρόπος για να επιτευχθεί αυτό είναι με την εφαρμογή του Bayesian ταξινομητή. Είδαμε, όμως, ότι η εφαρμογή του βέλτιστου ταξινομητή δεν είναι πάντα εύκολη, λόγω έλλειψης πληροφορίας σχετικά με τις αναγκαίες για τους υπολογισμούς pdf's.

Στο κεφάλαιο αυτό θα "ξεχάσουμε" τον κατά Bayes ταξινομητή και θα προσπαθήσουμε να διαιρέσουμε το χώρο με άλλους τρόπους. Οι τρόποι αυτοί θα είναι βέλτιστοι ως προς κάποιο κριτήριο, το οποίο όμως δεν είναι κατ' ανάγκη η πιθανότητα λάθους. Οι ταξινομητές που θα εστιάσουμε έχουν τις ρίζες τους στην έρευνα που άρχισε το τέλος της δεκαετίας του '50 με στόχο να κατασκευαστούν "μηχανές που μαθαίνουν". Στην κατεύθυνση αυτή έγινε προσπάθεια να μοντελοποιηθεί ο τρόπος με τον οποίο λειτουργούν τα βασικά στοιχεία του ανθρώπινου εγκεφάλου, οι νευρώνες. Αυτός είναι και ο λόγος που οι ταξινομητές που θα εξετάσουμε στο κεφάλαιο αυτό είναι γνωστοί και ως νευρωνικά δίκτυα.

Στην αρχή του κεφαλαίου, στην Ενότητα 8.1, θα εστιάσουμε στο βασικό δομικό στοιχείο τέτοιων ταξινομητών, που είναι γνωστό ως Perceptron ή νευρώνας. Ο απλός αυτός ταξινομητής διαχωρίζει το χώρο γραμμικά με ένα υπερεπίπεδο. Στη συνέχεια, Ενότητα 8.2, το βασικό αυτό δομικό στοιχείο χρησιμοποιείται για την ανάπτυξη πιο πολύπλοκων αρχιτεκτονικών, με δυνατότητες μη γραμμικής διαίρεσης του χώρου. Στην ίδια ενότητα γίνεται η περιγραφή της μεθοδολογίας εκπαίδευσης πολυστρωματικών Perceptrons.

ΕΝΟΤΗΤΑ 8.1 Ο ΑΛΓΟΡΙΘΜΟΣ PERCEPTRON

Στην ενότητα αυτή θα εστιάσουμε και πάλι στο απλούστερο πρόβλημα των δύο κλάσεων ω_1 και ω_2. Επιπλέον, θα υποθέσουμε ότι οι δύο αυτές κλάσεις είναι γραμμικά διαχωρίσιμες. Αυτό σημαίνει, ότι υπάρχει τουλάχιστον ένα υπερεπίπεδο που να διαχωρίζει πλήρως τις δύο κλάσεις. Με άλλα λόγια, υπάρχουν συντελεστές $w_0^*, w_1^*, ..., w_l^*$, τέτοιοι



Σχήμα 8.1: *Γεωμετρική ερμηνεία των παραμέτρων που περιγράφουν μία ευθεία,* $\mathbf{w}^T\mathbf{x}+w_0=0$, *για* $w_1>0$, $w_2>0$, $w_0<0$.

ώστε το υπερεπίπεδο με εξίσωση $w_1^* x_1 + w_2^* x_2 + ... + w_l^* x_l + w_0^* = 0$, ή $\mathbf{x}^T \mathbf{w}^* + w_0^* = 0$ να διαχωρίζει πλήρως τις δύο κλάσεις.

Στο Σχ. 8.1 φαίνεται η γεωμετρική ερμηνεία (για την περίπτωση δύο διαστάσεων) των παραμέτρων που υπεισέρχονται στη γενική εξίσωση:

$$g(\mathbf{x})=\mathbf{w}^T\mathbf{x}+w_0=0 \tag{8.1}$$

Για l διαστάσεις αυτή είναι η εξίσωση ενός υπερεπιπέδου, στον l-διάστατο χώρο, των χαρακτηριστικών διανυσμάτων. Από τη μια πλευρά (+) του υπερεπιπέδου έχουμε $g(\mathbf{x})>0$ και από την άλλη πλευρά (-) $g(\mathbf{x})<0$. Το διάνυσμα παραμέτρων $\mathbf{w}$ είναι κάθετο στο υπερεπίπεδο $g(\mathbf{x})=0$. Οι αποστάσεις d και z του σχήματος αποδεικνύεται ότι δίνονται από τις σχέσεις:

$$d = \frac{|w_0|}{\sqrt{w_1^2 + w_2^2}} \qquad \text{και} \qquad z = \frac{|g(\mathbf{x})|}{\sqrt{w_1^2 + w_2^2}}$$

Όταν $w_0=0$, το υπερεπίπεδο διέρχεται από την αρχή των αξόνων. Για τους σκοπούς του κεφαλαίου μπορούμε να γράψουμε την (8.1) ισοδύναμα ως

$$\hat{\mathbf{x}}^T \hat{\mathbf{w}} = 0 \qquad 8.2$$

όπου $\hat{\mathbf{x}}^T=(\mathbf{x}^T,1)$ και $\hat{\mathbf{w}}^T=(\mathbf{w}^T,w_0)$. Με άλλα λόγια, η (8.2) μας λέει ότι μπορούμε να εργαζόμαστε ισοδύναμα στον (l+1)-διάστατο χώρο και με υπερεπίπεδα που διέρχονται από την αρχή των αξόνων. Στη συνέχεια θα εργαζόμαστε με υπερεπίπεδα της μορφής $\mathbf{w}^T\mathbf{x}=0$ χωρίς, επομένως, βλάβη της γενικότητας.

Το ζητούμενο τώρα είναι, εάν μας δοθούν τα N διανύσματα εκπαίδευσης του **X**, πώς μπορούμε να προσδιορίσουμε ένα υπερεπίπεδο που να διαχωρίζει τις κλάσεις. Τη λύση την έδωσε πρώτος ο Rosenblatt [1]. Η πορεία που θα ακολουθήσουμε είναι αυτή της ελαχιστοποίησης ενός κατάλληλα επιλεγμένου κόστους με τη βοήθεια ενός αναδρομικού αλγορίθμου. Η συνάρτηση κόστους, που υιοθετείται για το παρόν πρόβλημα, έχει ενδιαφέρον. Εάν $\mathbf{w}$ είναι το διάνυσμα των συντελεστών ενός υπερεπιπέδου, ορίζουμε τη **Perceptron συνάρτηση** κόστους ως

$$J(\mathbf{w}) = \sum_{\mathbf{x}\in Y} (\mathbf{x}^T\mathbf{w})\delta_{\mathbf{x}} \tag{8.3}$$

όπου Y το σύνολο εκείνων των $\mathbf{x} \in X$ που ταξινομούνται λάθος από το $\mathbf{w}$ και

$$\delta_{\mathbf{x}} = \begin{cases} -1, & \mathbf{x} \in \omega_1 \\ +1, & \mathbf{x} \in \omega_2 \end{cases} \tag{8.4}$$

Προφανώς $J(\mathbf{w}) \geq 0$. Πράγματι, εάν $\mathbf{x} \in \omega_1$ και ταξινομείται λάθος, τότε θα πρέπει

$\mathbf{x}^T\mathbf{w} < 0$, που όμως πολλαπλασιάζεται με -1 και γίνεται θετικό. Με ανάλογο σκεπτικό καταλήγουμε ότι η ποσότητα είναι θετική όταν $\mathbf{x} \in \omega_2$ και ταξινομείται λάθος. Πολλαπλασιάζοντας επομένως με το κατάλληλο κάθε φορά $\delta_{\mathbf{x}}$ καθιστούμε το γινόμενο πάντα θετικό. Η J(**w**) μηδενίζεται όταν όλα τα $\mathbf{x} \in X$ ταξινομούνται σωστά και Y είναι το κενό σύνολο. Ένας αναδρομικός αλγόριθμος που ελαχιστοποιεί τη συνάρτηση κόστους (8.3) είναι ο **αλγόριθμος Perceptron**, το t-βήμα του οποίου δίνεται από:

$$\mathbf{w}(t+1) = \mathbf{w}(t) - \rho_t \sum_{x \in Y} \delta_{\mathbf{x}} \mathbf{x} \tag{8.5}$$

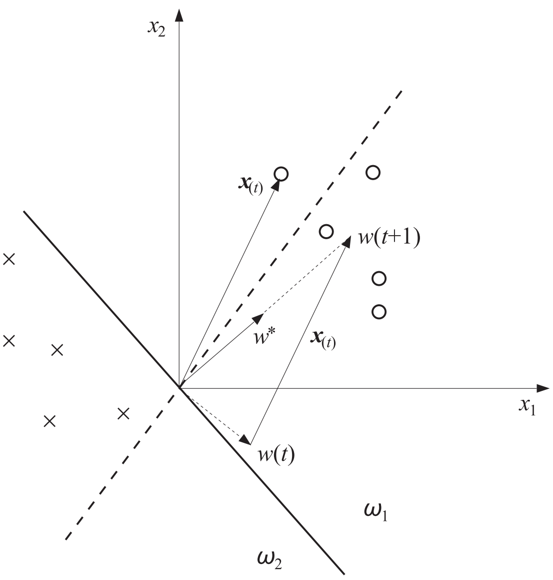
Για αρχικό όρο **w**(0) της ακολουθίας των διανυσμάτων-συντελεστών, που υπολογίζεται μέσω της (8.5) χρησιμοποιείται ένα τυχαίο μη μηδενικό διάνυσμα, και για ρ_t μια κατάλληλα επιλεγμένη ακολουθία, π.χ., ρ_t=σταθερά<2, ή ρ_t=σταθερά/t. Αποδεικνύεται ότι, κάτω από αυτές τις προϋποθέσεις ο αλγόριθμος Perceptron συγκλίνει σε μια λύση μετά από **πεπερασμένο αριθμό βημάτων αναδρομής**. Το πόσα βήματα απαιτούνται εξαρτάται από την αρχική τιμή **w**(0) και την επιλογή της ρ_t. Για πιο πολλές λεπτομέρειες και μερικές θεωρητικές αναζητήσεις ο αναγνώστης παραπέμπεται στο [2]. Θα πρέπει να τονιστεί ότι, ο αλγόριθμος δεν συγκλίνει όταν οι κλάσεις δεν είναι γραμμικά διαχωρίσιμες. Η αναδρομή (8.5) μας λέει ότι στο βήμα t, όπου έχουμε το **w**(t), υπολογίζουμε τα εσωτερικά γινόμενα $\mathbf{x}^T\mathbf{w}(t)$ με όλα τα $\mathbf{x} \in X$, και δημιουργούμε το άθροισμα στην (8.5) από εκείνα τα διανύσματα που ταξινομούνται λάθος. Στη συνέχεια κάνουμε τη διόρθωση **w**(t+1).

Η (8.5) δεν είναι η μοναδική μορφή του αλγόριθμου Perceptron. Μία άλλη δημοφιλής μορφή είναι η ακόλουθη:

$$\begin{aligned} \mathbf{w}(t+1) &= \mathbf{w}(t) + \rho\mathbf{x}_{(t)} \quad \text{εάν} \quad \mathbf{x}_{(t)} \in \omega_1 \quad \text{και} \quad \mathbf{w}^T(t)\mathbf{x}_{(t)} \le 0 \\ \mathbf{w}(t+1) &= \mathbf{w}(t) - \rho\mathbf{x}_{(t)} \quad \text{εάν} \quad \mathbf{x}_{(t)} \in \omega_2 \quad \text{και} \quad \mathbf{w}^T(t)\mathbf{x}_{(t)} \ge 0 \\ \mathbf{w}(t+1) &= \mathbf{w}(t) \qquad \text{διαφορετικά} \end{aligned} \tag{8.6}$$

Στη μορφή (8.6) παρουσιάζονται τα διανύσματα εκπαίδευσης στον αλγόριθμο το ένα μετά το άλλο και χρησιμοποιείται ένα μόνο διάνυσμα εκπαίδευσης σε κάθε βήμα του αλγορίθμου. Με $\mathbf{x}_{(t)}$ συμβολίζεται το δείγμα που έχει σειρά να παρουσιαστεί στο βήμα t. Από την ίδια σχέση φαίνεται ότι αν το δείγμα αυτό ταξινομείται σωστά, δεν γίνεται διόρθωση. Εάν ταξινομείται λάθος, τότε επιβάλουμε στο **w**(t) διορθωτικές κινήσεις. Στο Σχ. 8.2, φαίνεται πώς ο αλγόριθμος στρέφει το επίπεδο, ώστε να τοποθετήσει το $\mathbf{x}_{(t)}$ στη σωστή κλάση. Η επιλογή του ρ εδώ παίζει προφανώς σημαντικό ρόλο. Αποδεικνύεται ότι και αυτός ο αλγόριθμος συγκλίνει σε πεπερασμένο αριθμό βημάτων, υπό την προϋπόθεση βέβαια ότι οι κλάσεις είναι γραμμικά διαχωρίσιμες.

Πρέπει να τονιστεί ότι για να επιτευχθεί σύγκλιση θα πρέπει, συνήθως, να χρησιμοποιήσουμε τα διανύσματα του X περισσότερες από μία φορές. Κάθε φορά που εξαντλούμε τα στοιχεία του X λέμε ότι συμπληρώσαμε μία **εποχή** (*epoch*). Στη συνέχεια παρουσιάζονται τα διανύσματα στον αλγόριθμο από την αρχή. Για να επιτευχθεί σύγκλιση απαιτείται ένας αριθμός εποχών.



Σχήμα 8.2: *Η ευθεία* ***w****(t)(διακεκομένη) ταξινομεί λάθος το* $\mathbf{x}_{(t)}$*. Ο αλγόριθμος Perceptron στρέφει το επίπεδο* ($\mathbf{w}$(t+1)) *ώστε να περιλάβει το* $\mathbf{x}_{(t)}$ *στη σωστή κλάση.*

Μετά τη σύγκλιση του αλγορίθμου Perceptron και τον υπολογισμό των παραμέτρων w_i, i=0,…,*l*, η ταξινόμηση γίνεται σύμφωνα με το πρόσημο του $\mathbf{w}^T\mathbf{x}$, όπου **x** το (επαυξημένο με 1) προς ταξινόμηση άγνωστο διάνυσμα. Όταν το πρόσημο της ποσότητας είναι θετικό το **x** βρίσκεται στη μια πλευρά του υπερεπιπέδου και ταξινομείται στην ω_1, ενώ όταν το πρόσημο είναι αρνητικό το **x** βρίσκεται στην άλλη πλευρά του υπερεπιπέδου και ταξινομείται στην ω_2.

Παράδειγμα 1/Κεφάλαιο 8

Σε ένα πρόβλημα ταξινόμησης δύο κλάσεων χρησιμοποιούνται διανύσματα χαρακτηριστικών με δύο συνιστώσες. Στο Σχ. 8.3 έχουν σχεδιαστεί με ‘+’ και ‘o’ τα αναγνωρισμένα διανύσματα καθώς και η ευθεία με εξίσωση $x_1+x_2-0.5$ (αυτή με τη διακεκομμένη γραμμή), η οποία έχει προκύψει κατά την εκτέλεση του t βήματος της (8.5). Όπως φαίνεται από το σχήμα, η ευθεία αυτή χωρίζει το χώρο των χαρακτηριστικών αφήνοντας όμως τα διανύσματα $(0.4,0.05)^T$ και $(-0.2,0.75)^T$ σε λάθος περιοχή. Να εκτελεστεί η επόμενη επανάληψη του αλγορίθμου της (8.5), να υπολογιστεί η νέα εξίσωση της ευθείας, να σχεδιαστεί η ευθεία στο διάγραμμα και να εξεταστεί αν διαχωρίζει σωστά όλα τα δείγματα.

Δίνεται $\rho_t=\rho=0.7$.

Απάντηση

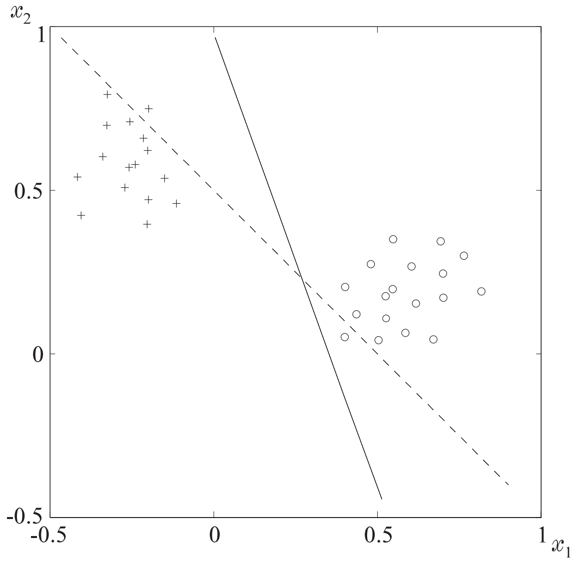
Από τα δεδομένα προκύπτουν: $\mathbf{w}(t)=(1,1,-0.5)^T$, $\mathbf{x}_1=(0.4,0.05,1)^T$, $\mathbf{x}_2=(-0.2,0.75,1)^T$, $\delta_{\mathbf{x1}}=-1$, και $\delta_{\mathbf{x2}}=1$.

Η επόμενη επανάληψη, σύμφωνα με την (8.5) δίνει:

$\mathbf{w}(t+1)= \mathbf{w}(t)-\rho(\delta_{x1}\mathbf{x}1+ \delta_{x2}\mathbf{x}_2)$, ή

$\mathbf{w}(t+1)= (1,1,-0.5)^T-0.7[(-1)\ (0.4,0.05,1)^T+(+1)\ (-0.2,0.75,1)^T]$, ή

$\mathbf{w}(t+1)= (1.42,0.51,-0.5)^T$.



Σχήμα 8.3 *Ένα παράδειγμα του αλγόριθμου Perceptron.*

Από το διάνυσμα των συντελεστών προκύπτει η εξίσωση της ευθείας $1.42x_1+0.51x_2-0.5=0$, η οποία όταν χαραχθεί στο διάγραμμα του σχήματος (συνεχής γραμμή) διαπιστώνουμε ότι ταξινομεί σωστά όλα τα διανύσματα και, επομένως, η (t+1) επανάληψη είναι και η τελική.

Άσκηση Αυτοαξιολόγησης 1/Κεφάλαιο 8.

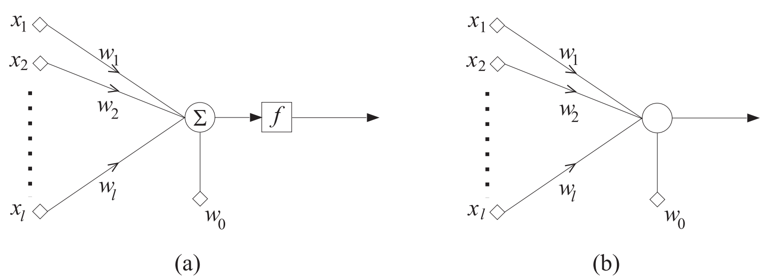
Η κλάση ω_1 περιλαμβάνει τα διανύσματα $(0,0)^T$ και $(0,1)^T$, και η κλάση ω_2 τα $(1,1)^T$ και $(1,0)^T$. Να χρησιμοποιηθεί ο αλγόριθμος Perceptron (8.6) με $\rho=1$ και επαυξημένου $\mathbf{w}(0)=(0,1,0)^T$ ώστε να υπολογίσει μια ευθεία που να διαχωρίζει τις κλάσεις.

8.1.1 Η ΑΡΧΙΤΕΚΤΟΝΙΚΗ PERCEPTRON

Στο Σχ. 8.4 φαίνεται η αρχιτεκτονική Perceptron ή όπως επίσης λέγεται ενός **νευρώνα**. Τα ουσιώδη χαρακτηριστικά $x_1,\ldots,x_l$ διεγείρουν τα ***στοιχεία*** *(nodes)* της εισόδου. Καθένα από αυτά πολλαπλασιάζεται με το αντίστοιχο βάρος w_i. Τα w_i είναι γνωστά ως **συνάψεις**. Τα επιμέρους γινόμενα αθροίζονται μαζί με το w_0, γνωστό ως **κατώφλιο** (*threshold*). Ο υπολογισμός των συνάψεων και του κατωφλίου γίνεται με τη βοήθεια του αλγόριθμου Perceptron. Στη συνέχεια η έξοδος του αθροιστή περνά από μία μη γραμμική συνάρτηση f(.), που στην προκειμένη περίπτωση είναι η μοναδιαία βηματική συνάρτηση.

$$f(z)=\begin{cases}1, & z>0\\ 0, & z<0\end{cases}$$

Η συνάρτηση αυτή είναι γνωστή και ως **συνάρτηση ενεργοποίησης** (*activation*) του Perceptron. Με άλλα λόγια, το Perceptron υλοποιεί ένα υπερεπίπεδο $\mathbf{w}^T\mathbf{x}+w_0=0$ και η έξοδος είναι 1 ή 0 ανάλογα από το αν το $\mathbf{x}$ βρίσκεται στη μία ή την άλλη πλευρά του υπερεπιπέδου. Σε άλλες περιπτώσεις η f(.) μπορεί να έχει και άλλη μορφή. Για παράδειγμα, η έξοδος μπορεί να είναι -1 ή 1.



Σχήμα 8.4: *(a) Αρχιτεκτονική Perceptron και (b) συμβολισμός όπου ο αθροιστής και η συνάρτηση ενεργοποίησης συμβολίζονται από κοινού με ένα κύκλο.*

Το βασικό αυτό στοιχείο χρησιμοποιήθηκε από τον Rosenblat στα τέλη της δεκαετίας του 50 για να μοντελοποιήσει τους βασικούς νευρώνες του εγκεφάλου. Είναι ένα απλό παράδειγμα μιας **μηχανής που μαθαίνει** (*learning mahine*). Με άλλα λόγια, ενός στοιχείου του οποίου οι ελεύθερες παράμετροι (π.χ. συνάψεις κατώφλι) υπολογίζονται με τη βοήθεια ενός αλγορίθμου εκμάθησης (π.χ., αλγόριθμος Perceptron) που ενεργεί πάνω στο σύνολο των διανυσμάτων εκπαίδευσης.

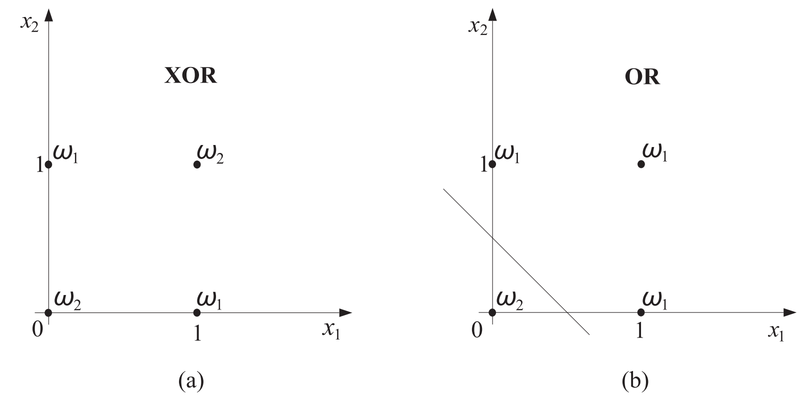
ΕΝΟΤΗΤΑ 8.2 ΜΗ ΓΡΑΜΜΙΚΑ ΔΙΑΧΩΡΙΣΙΜΑ ΠΡΟΒΛΗΜΑΤΑ

8.2.1 PERCEPTRONS 2-ΣΤΡΩΜΑΤΩΝ

Το Perceptron είναι ένας βασικός ταξινομητής κατάλληλος για γραμμικά διαχωρίσιμα προβλήματα. Τα προβλήματα όμως αυτά αποτελούν μειοψηφία στην πράξη. Στη συνέχεια θα εστιάσουμε σε ένα, γνωστό σε όλους μας, πρόβλημα που δεν επιδέχεται γραμμική προσέγγιση.

Οι Boolean συναρτήσεις μπορούν να εκληφθούν ότι εκτελούν χρέη ταξινόμησης σε πρόβλημα δύο κλάσεων. Πράγματι, η έξοδος μιας τέτοιας συνάρτησης είναι ‘1’, ή ‘0’ εξαρτώμενη από την τιμή των δυαδικών δεδομένων $\mathbf{x}=(x_1, x_2,..., x_l)^T$ της εισόδου της.

Με αυτό τον τρόπο το **x** ταξινομείται στην κλάση ω_1 (‘1’), ή ω_2 (‘0’). Η Boolean XOR πράξη είναι ένα κλασικό μη γραμμικό πρόβλημα ταξινόμησης. Στον Πίνακα 8.1 δίνονται οι Αληθοπίνακες των λογικών πράξεων XOR και OR για δύο εισόδους x_1, x_2. Στα Σχήματα



Σχήμα 8.5: *Γεωμετρία του προβλήματος (a) XOR και (b) OR.*

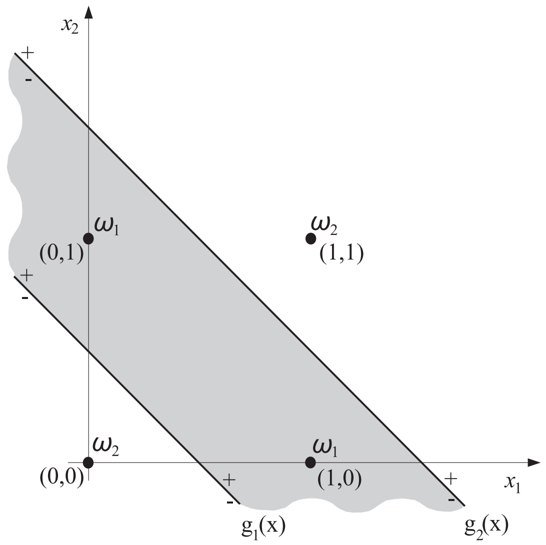
8.5.α και 8.5.β φαίνονται τα αντίστοιχα σημεία στο δισδιάστατο χώρο χαρακτηριστικών μαζί με τις κλάσεις που ανήκουν. Είναι σαφές ότι δεν υπάρχει μια μοναδική ευθεία (υπερεπίπεδο) που να χωρίζει τις δύο κλάσεις στο XOR πρόβλημα, ενώ αυτό είναι εφικτό στο γραμμικά διαχωρίσιμο πρόβλημα της πράξης OR.

x_1	x_2	XOR	Κλάση
0	0	0	ω_2
0	1	1	ω_1
1	0	1	ω_1
1	1	0	ω_2

x_1	x_2	OR	Κλάση
0	0	0	ω_2
0	1	1	ω_1
1	0	1	ω_1
1	1	1	ω_1

Πίνακας 8.1 *Οι Αληθοπίνακες των πυλών XOR και OR.*

Για την αντιμετώπιση του προβλήματος XOR θα ακολουθήσουμε την εξής πορεία. Θα σχηματίσουμε δύο αντί μιας ευθείας (υπερεπίπεδα), όπως φαίνεται στο Σχ. 8.6. Η κλάση ω_1 βρίσκεται μεταξύ των ευθειών και η ω_2 εκτός των ευθειών. Ας δούμε τώρα αυτή τη διαδικασία λίγο διαφορετικά, που θα μας οδηγήσει αργότερα σε γενικεύσεις. Στην ουσία χωρίζουμε τους υπολογισμούς μας σε δύο φάσεις. Στην πρώτη φάση δημιουργούμε δύο υπερεπίπεδα $g_1(\mathbf{x})=0$ και $g_2(\mathbf{x})=0$, χρησιμοποιώντας δύο Perceptrons. Η έξοδος των Perceptrons αυτών είναι y_1 και y_2 αντίστοιχα, όπως φαίνεται και στον Πίνακα 8.2. Πράγματι, σε σχέση με το πρώτο υπερεπίπεδο (ευθεία) $g_1(\mathbf{x})$ του Σχ. 8.6 (έξοδος y_1 στον

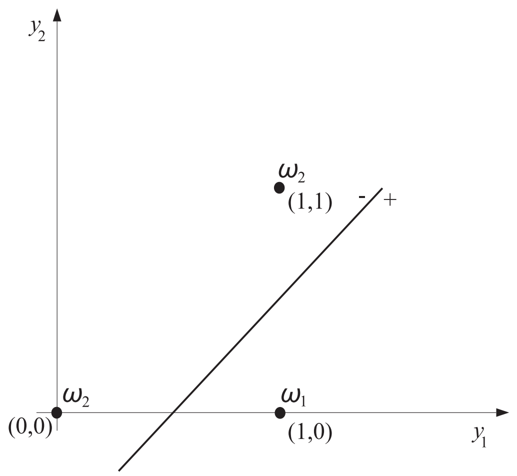


Σχήμα 8.6: *Ταξινόμηση του XOR με δύο ευθείες.*

στον Πίνακα) μόνο το (0,0) σημείο βρίσκεται στη (-) πλευρά και όλα τα άλλα στην (+). Το τοπίο αλλάζει για το υπερεπίπεδο (ευθεία) $g_2(\mathbf{x})$ (έξοδος y_2). Συνδυάζοντας τώρα τα

1η φάση				2η φάση
x_1	x_2	y_1	y_2	
0	0	0(-)	0(-)	ω_2 (0)
1	0	1(+)	0(-)	ω_1 (1)
0	1	1(+)	0(-)	ω_1 (1)
1	1	1(+)	1(+)	ω_2 (0)

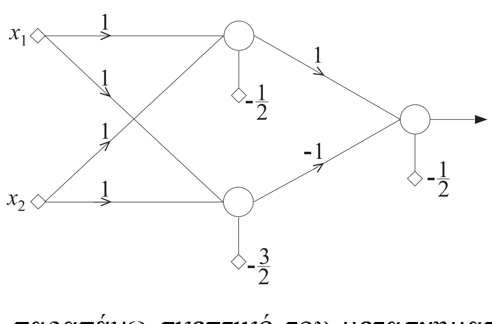
Πίνακας 8.2 *Αληθοπίνακας για τις δύο φάσεις υπολογισμού του XOR προβλήματος.*



Σχήμα 8.7: *Τα σημεία του XOR προβλήματος είναι γραμμικά διαχωρίσιμα στον νέο χώρο (y_1,y_2).*

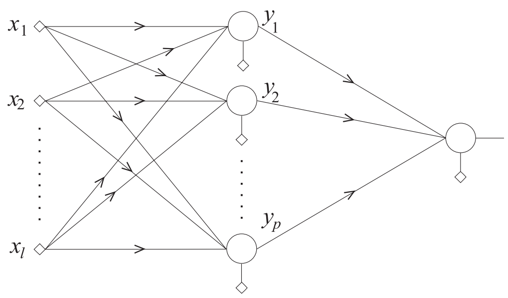
y_1, y_2, Σχ.8.7, παρατηρούμε ότι τα σημεία της κλάσης ω_2 (μεταξύ των ευθειών στο Σχ. 8.6) αντιστοιχούν σε ένα σημείο (1, 0) και τα εκτός των ευθειών (Σχ. 8.6) στα σημεία (0, 0) και (1, 1). Τα τρία όμως αυτά σημεία είναι γραμμικά διαχωρίσιμα!

Με άλλα λόγια, στην πρώτη φάση υπολογισμών μετασχηματίζουμε το χώρο των χαρακτηριστικών σε ένα νέο χώρο (y_1, y_2). Στον μετασχηματισμένο αυτό χώρο, οι κλάσεις είναι γραμμικά διαχωρίσιμες και στη δεύτερη φάση υπολογισμών οι κλάσεις διαχωρίζονται με ένα τρίτο Perceptron στο χώρο (y_1, y_2). Το Σχ 8.8 δείχνει μια τέτοια αρχιτεκτονική που υλοποιεί το XOR πρόβλημα. Το δίκτυο που προκύπτει είναι γνωστό και ως Perceptron 2 στρωμάτων. Τα ουσιώδη χαρακτηριστικά εφαρμόζονται στα στοιχεία εισόδου. Τα δύο Perceptrons ή νευρώνες του πρώτου στρώματος, γνωστό και ως **κρυφό στρώμα**, μετασχηματίζουν την είσοδο στο χώρο (y_1, y_2), που είναι ο συνδυασμός των εξόδων των Perceptrons του πρώτου στρώματος. Στη συνέχεια, το τρίτο Perceptron, γνωστό ως **Perceptron** ή **νευρώνας εξόδου** υλοποιεί το υπερεπίπεδο που λύνει το γραμμικό πρόβλημα στον (y_1, y_2) χώρο. Τα βάρη του Σχ. 8.8 αντιστοιχούν στις ευθείες των σχημάτων 8.6 και 8.7. Η αρχιτεκτονική αυτή γενικεύεται άμεσα και για *l* εισόδους και περισσότερους από δύο, p, νευρώνες στο κρυφό στρώμα, Σχ.
8.9.



Σχήμα 8.8: *Perceptron 2 στρωμάτων για την επίλυση του XOR προβλήματος.*

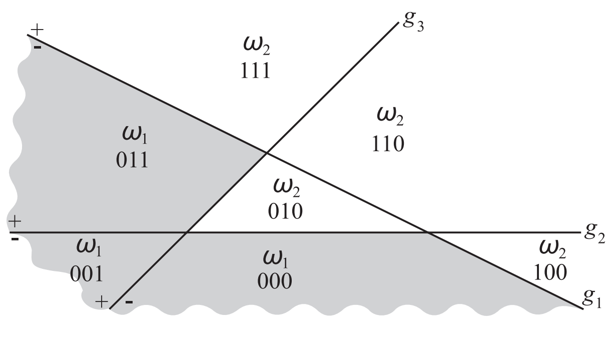
Το παραπάνω σκεπτικό του μετασχηματισμού του χώρου των χαρακτηριστικών σε έναν άλλο χώρο δεν είναι μεμονωμένο, αλλά εντάσσεται σε ένα γενικότερο σημαντικό θεωρητικό αποτέλεσμα. Αποδεικνύεται ότι μεταβαίνοντας σε ένα νέο χώρο, εν γένει μεγαλύτερης διάστασης, η πιθανότητα το πρόβλημα να μετατραπεί σε γραμμικά διαχωρίσιμο αυξάνει [2].



Σχήμα 8.9: *Αρχιτεκτονική Perceptron 2 στρωμάτων*

8.2.2 ΔΙΑΧΩΡΙΣΤΙΚΗ ΙΚΑΝΟΤΗΤΑ ΤΩΝ PERCEPTRONS ΔΥΟ ΣΤΡΩΜΑΤΩΝ

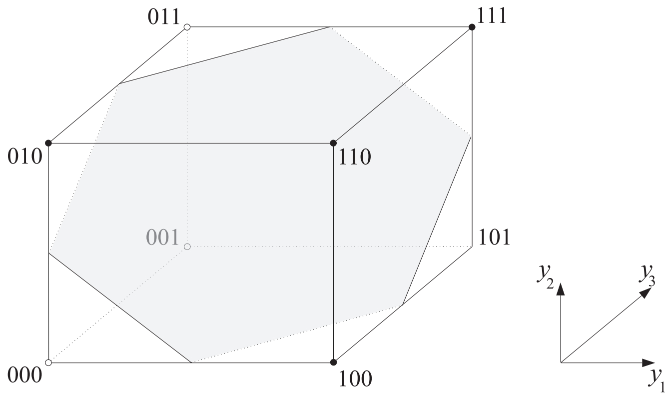
Στο Σχ. 8.9, καθένας από τους p νευρώνες του κρυφού στρώματος υλοποιεί ένα υπερεπίπεδο, στον *l*-διάστατο χώρο των χαρακτηριστικών. Η έξοδος του κάθε νευρώνα είναι 1 ή 0 ανάλογα με τη θέση του διανύσματος εισόδου ως προς το αντίστοιχο υπερεπίπεδο. Το Σχ. 8.10 αντιστοιχεί σε ένα Perceptron στο δισδιάστατο (*l*=2) χώρο με τρεις νευρώνες στο κρυφό στρώμα (ένας για κάθε ευθεία). Οι τρεις ευθείες g_1, g_2, g_3 χωρίζουν το επίπεδο των χαρακτηριστικών σε επτά περιοχές (πολύεδρα στη γενική περίπτωση). Σε καθεμιά από αυτές τις περιοχές αντιστοιχεί και μια τριάδα αριθμών (y_1, y_2, y_3) που δείχνει τη σχετική θέση (1(+) και 0(-)) της περιοχής ως προς τις ευθείες g_1, g_2, g_3, αντίστοιχα. Στο σχήμα επίσης οι δύο περιοχές έχουν αντιστοιχηθεί, στις δύο κλάσεις.



Σχήμα 8.10: *Περιοχές που χωρίζουν τον δισδιάστατο χώρο τρεις ευθείες. Η κάθε τριάδα των ακεραίων δείχνει τη σχετική θέση (+ ή -) της κάθε περιοχής ως προς τις τρεις ευθείες.*

Με άλλα λόγια, ο "κωδικός" κάθε περιοχής είναι το σημείο (y_1, y_2, y_3) στο χώρο που μετασχηματίζουν την 2-διάστατη είσοδο οι τρεις νευρώνες του κρυφού στρώματος. *Ουσιαστικά, αυτός είναι ένας μετασχηματισμός της εισόδου στις κορυφές ενός μοναδιαίου κύβου* (υπερκύβου στην περίπτωση p>3 νευρώνων). Το κρίσιμο ζήτημα εδώ είναι το πώς κατανέμονται οι περιοχές αυτές στις δύο κλάσεις ω_1, ω_2. Για το παράδειγμα του σχήματος

8.10, οι κλάσεις είναι έτσι κατανεμημένες, ώστε οι κορυφές του κύβου που αντιστοιχούν στην κλάση ω_1 να είναι γραμμικά διαχωρίσιμες από τις κορυφές που αντιστοιχούν στην κλάση ω_2. (Παρατηρήστε ότι μία κορυφή, η 101 γι' αυτή την περίπτωση δεν αντιστοιχεί σε περιοχή. Αυτό συμβαίνει πάντα και η κορυφή αυτή δεν επηρεάζει το πρόβλημα.) Έτσι, η υλοποίηση ενός επιπέδου από το νευρώνα εξόδου διαχωρίζει τις κλάσεις, Σχ. 8.11. Εάν όμως οι περιοχές κατανέμονται στις κλάσεις έτσι ώστε οι κορυφές του κύβου που αντιστοιχούν στην κλάση ω_1 να μην είναι γραμμικά διαχωρίσιμες από εκείνες που αντιστοιχούν στην ω_2, τότε δεν είναι δυνατόν το επίπεδο που υλοποιεί ο νευρώνας εξόδου να διαχωρίσει τις δύο κλάσεις. Για παράδειγμα, εάν η ω_1 αποτελείται από τις περιοχές 000 και 111 και η ω_2 από τις υπόλοιπες, δεν θα ήταν δυνατόν να σχεδιάσουμε ένα επίπεδο που να έχει στη μια του πλευρά τις κορυφές της ω_1 και στην άλλη πλευρά τις υπόλοιπες.



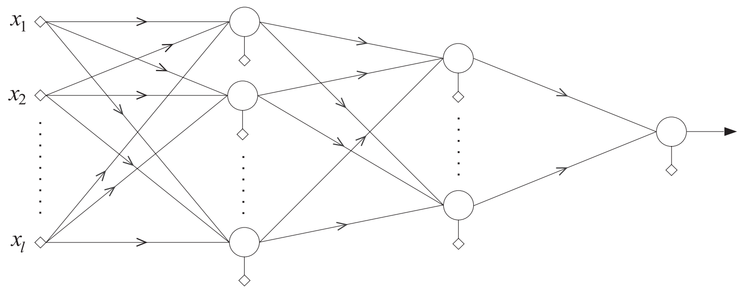
Σχήμα 8.11: *Οι κωδικοί των περιοχών του σχήματος 8.10 αντιστοιχούν στις κορυφές ενός κύβου. Το επίπεδο, που υλοποιεί ο νευρώνας εξόδου, διαχωρίζει τις κλάσεις.*

Συνοψίζοντας, μπορούμε να πούμε ότι ένα Perceptron 2-στρωμάτων είναι δυνατόν να διαχωρίσει δύο κλάσεις που αποτελούνται από ενώσεις πολυεδρικών περιοχών, αλλά όχι οποιαδήποτε ένωση πολυεδρικών περιοχών.

8.2.3 PERCEPTRONS 3-ΣΤΡΩΜΑΤΩΝ

Το προηγούμενο "αδιέξοδο" που δημιουργήθηκε με τα Perceptron 2-στρωμάτων μπορούμε εύκολα να το υπερκεράσουμε χρησιμοποιώντας, για μια ακόμη φορά, την εμπειρία με την οποία μας εμπλούτισε το XOR πρόβλημα. Εάν οι δύο κλάσεις αντιστοιχούν στις κορυφές του υπερκύβου (μετά τον αρχικό μετασχηματισμό) με τρόπο που να μην είναι γραμμικά διαχωρίσιμες, τότε ας επιχειρήσουμε να σχεδιάσουμε περισσότερα από ένα υπερεπίπεδα και να δημιουργήσουμε ένα ακόμη στρώμα στο Perceptron. Η λύση εδώ περιλαμβάνει τρεις φάσεις.

Στην πρώτη φάση, ο χώρος των χαρακτηριστικών μετασχηματίζεται στις κορυφές ενός υπερκύβου, με τα Perceptrons του 1ου στρώματος. Στη συνέχεια, επιλέγουμε όλες τις κορυφές που αντιστοιχούν στη μία από τις δύο κλάσεις, ας πούμε την ω_1. Για καθεμιά από τις κορυφές αυτές, υλοποιούμε (στο μετασχηματισμένο χώρο) ένα υπερεπίπεδο με τη βοήθεια ενός νευρώνα. Οι συντελεστές υπολογίζονται έτσι ώστε η συγκεκριμένη κορυφή να βρίσκεται στην (+) πλευρά και *όλες* οι άλλες στην (-). Άρα η έξοδος του νευρώνα αυτού είναι 1 μόνο για τα σημεία της περιοχής που αντιστοιχεί στη συγκεκριμένη κορυφή και 0 για όλες τις άλλες περιοχές (ανεξάρτητα της κλάσης). Επαναλαμβάνουμε την παραπάνω διαδικασία ξεχωριστά για καθεμιά από τις κορυφές που αντιστοιχούν στην κλάση ω_1. Εάν, για παράδειγμα, είχαμε p τέτοιες κορυφές, στη δεύτερη αυτή φάση υπολογισμών θα χρειαζόμαστε p νευρώνες. Παρατηρήστε ότι, κάθε φορά, που εμφανίζεται ένα **x** από κάποια περιοχή που ανήκει στην ω_1, τότε ο ένας από τους p νευρώνες θα δίνει έξοδο 1 και όλοι οι άλλοι 0. Αντίθετα, όταν το **x** ανήκει σε οποιαδήποτε από τις περιοχές της ω_2, τότε όλοι οι p νευρώνες θα δίνουν έξοδο 0. Επομένως, παρέχοντας τις εξόδους των p νευρώνων ως είσοδο σε ένα άλλο νευρώνα που υλοποιεί μια OR πύλη, η έξοδος του νευρώνα αυτού θα είναι πάντα 1 για την ω_1 και 0 για την ω_2.



Σχήμα 8.12: *Αρχιτεκτονική Perceptron 3 στρωμάτων.*

Η αρχιτεκτονική του σχήματος 8.13 υλοποιεί την παραπάνω αλληλουχία υπολογισμών. Το νευρωνικό αυτό δίκτυο είναι γνωστό **ως Perceptron 3-στρωμάτων**. Τα ουσιώδη χαρακτηριστικά εφαρμόζονται στα στοιχεία εισόδου. Οι νευρώνες του **1ου κρυφού στρώματος** υλοποιούν τα υπερεπίπεδα της πρώτης φάσης υπολογισμών. Οι νευρώνες του **2ου κρυφού στρώματος** αντιστοιχούν στη 2η φάση υπολογισμών και πραγματοποιούν το διαχωρισμό των περιοχών. **Ο νευρώνας εξόδου** αντιστοιχεί στην τρίτη φάση υπολογισμών, υλοποιεί μια πύλη OR και πραγματοποιεί το διαχωρισμό των κλάσεων.

Με την παραπάνω κατασκευαστική απόδειξη, αποδείξαμε ότι ένα Perceptron 3-στρωμάτων μπορεί να διαχωρίσει σε δύο κλάσεις *οποιαδήποτε ένωση πολυεδρικών περιοχών στο χώρο των χαρακτηριστικών διανυσμάτων.* Το αποτέλεσμα αυτό είναι πολύ σημαντικό και αναδεικνύει τις δυνατότητες ταξινόμησης αυτού του τύπου νευρωνικών δικτύων.

Παρατηρήσεις:

1. Εάν έχουμε περισσότερες των δύο κλάσεων, η φιλοσοφία που εκθέσαμε είναι άμεσα επεκτάσιμη, αρκεί να αυξήσουμε τους νευρώνες εξόδου και να αντιστοιχήσουμε ένα νευρώνα σε κάθε κλάση. Κάθε φορά που ένα πρότυπο **x** από μία κλάση εφαρμόζεται στο δίκτυο, ο αντίστοιχος νευρώνας εξόδου θα δίνει 1 και οι υπόλοιποι 0.
2. Μέχρις στιγμής, αναδείξαμε τις δυνατότητες ταξινόμησης των πολυστρωματικών Perceptrons. Στην πράξη, βέβαια, ο υπολογισμός των συνάψεων δεν είναι δυνατόν να γίνει αναλυτικά. Όπως και για το βασικό Perceptron, έτσι και για τα γενικευμένα αυτά νευρωνικά δίκτυα θα πρέπει να αναπτύξουμε αλγορίθμους εκμάθηνσης.
3. Ένα άλλο βασικό ερώτημα, που συνδέεται με τα πολυστρωματικά Perceptrons, είναι ο υπολογισμός του αριθμού των νευρώνων για κάθε στρώμα. Στην πράξη, όπου οι περιοχές των κλάσεων δεν είναι γνωστές εκ των προτέρων, αυτό δεν είναι εύκολο θέμα. Επίσης, στην πράξη οι διάφορες περιοχές δεν είναι απαραίτητα πολύεδρα. Η μελέτη και οι δυνατότητες των νευρωνικών δικτύων, σε σχέση με άλλους ταξινομητές, αποτελεί ακόμη ανοικτό θέμα έρευνας.

Παράδειγμα 2/Κεφάλαιο 8

Δίνονται τα παρακάτω διανύσματα και οι κλάσεις στις οποίες ανήκουν:

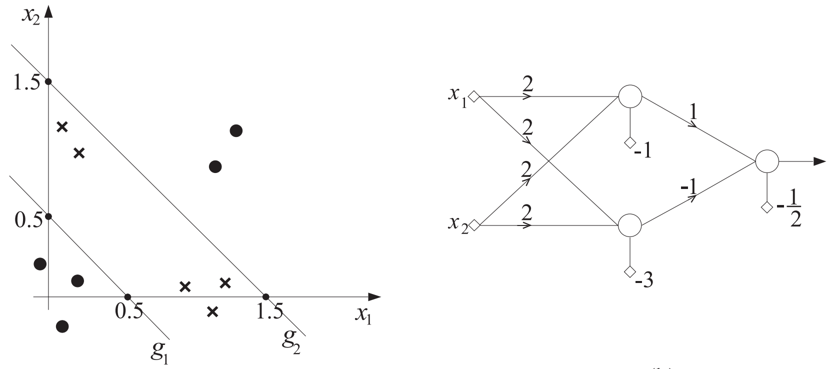
ω_1: $(0.1, -0.2)^T$, $(0.2, 0.1)^T$, $(-0.15, 0.2)^T$, $(1.1, 0.8)^T$, $(1.2, 1.1)^T$

ω_2: $(1.1, -0.1)^T$, $(1.25, 0.15)^T$, $(0.9, 0.1)^T$, $(0.1, 1.2)^T$, $(0.2, 0.9)^T$

Να ελεγχθεί εάν αυτά είναι γραμμικά διαχωρίσιμα, και να σχεδιαστεί μια κατάλληλη αρχιτεκτονική Perceptron που να τα διαχωρίζει.

Απάντηση

Από το Σχ. 8.13 φαίνεται πως τα διανύσματα αυτά δεν είναι γραμμικά διαχωρίσιμα. Δύο ευθείες γραμμές που διαχωρίζουν τις δύο κλάσεις είναι οι $2x_1+2x_2-1=0$ και $2x_1+2x_2-3=0$. Επομένως, το Perceptron δύο στρωμάτων του σχήματος 8.13 ταξινομεί τις κλάσεις. Ο νευρώνας εξόδου υλοποιεί την ευθεία $y_1 - y_2 - 1/2 = 0$ στον (y_1, y_2) χώρο.



(a) (b)

Σχήμα 8.13: *Γεωμετρία του παραδείγματος 8.2 και το Perceptron που επιλύει το πρόβλημα.*

Άσκηση Αυτοαξιολόγησης 2/Κεφαλαιου 8

Σχεδιάστε τις ακόλουθες τρεις ευθείες στο χώρο των δύο διαστάσεων: $x_1+x_2=0$, $x_2=1/4$, $-x_1+x_2=0$.

Για καθεμιά από τις περιοχές που δημιουργούνται από τις τομές τους, προσδιορίστε τις κορυφές του κύβου στις οποίες αντιστοιχούν, μετά το μετασχηματισμό που υλοποιούν οι νευρώνες του πρώτου στρώματος ενός πολυστρωματικού Perceptron. Συνδυάστε τις περιοχές που δημιουργούνται έτσι ώστε α) το πρόβλημα να είναι διαχωρίσιμο από Perceptron 2 στρωμάτων και β) από Perceptron 3 στρωμάτων. Για καθεμιά από τις περιπτώσεις προσδιορίστε αναλυτικά συνάψεις και παραμέτρους κατωφλίου.

8.2.4 Ο ΑΛΓΟΡΙΘΜΟΣ ΟΠΙΣΘΟΔΡΟΜΙΚΗΣ ΔΙΑΔΟΣΗΣ

Στην παρούσα υποενότητα θα περιγράψουμε μια μεθοδολογία εκπαίδευσης πολυστρωματικών Perceptrons. Υποθέτουμε ότι:

- Υπάρχουν N διανύσματα εκπαίδευσης, $X = \{\mathbf{x}(1), \mathbf{x}(2), ..., \mathbf{x}(N)\}$. Για καθένα από τα δείγματα αυτά γνωρίζουμε την κλάση που προέρχεται. Αυτό θα περιγράφεται με ένα αντίστοιχο διάνυσμα $\mathbf{y}(i) = (0,0,...,1,...,0)^T$, i=1, 2, …, N. Το διάνυσμα $\mathbf{y}(i) \in R^m$ έχει όλα τα στοιχεία 0, εκτός από το στοιχείο στην θέση r, όπου r η κλάση που ανήκει το αντίστοιχο **x**(i) και $r \in \{1,2,...,m\}$, για ένα πρόβλημα m κλάσεων.

- Θεωρούμε ότι έχουμε *l* στοιχεία εισόδου $(\mathbf{x} \in R^l)$, και m νευρώνες εξόδου, ένα για κάθε κλάση. Η έξοδος κάθε νευρώνα, όταν στην είσοδο εμφανίζεται το **x**(i), είναι $\hat{y}_p(i), p = 1,2,...,m$. Οι έξοδοι αυτοί αποτελούν τα στοιχεία ενός διανύσματος $\hat{\mathbf{y}}(i)$, το οποίο θα θέλαμε, ιδανικά, να ισούται με $\mathbf{y}(i)$. Στην πραγματικότητα όμως διαφέρουν. Αυτός είναι ο λόγος που το $\hat{\mathbf{y}}(i)$ είναι γνωστό ως το πραγματικό διάνυσμα εξόδου και το $\mathbf{y}(i)$ ως το επιθυμητό.

- Στην πράξη επιλέγουμε τη διαφορά μεταξύ επιθυμητών και πραγματικών εξόδων να είναι ελάχιστη. Ένα τέτοιο κριτήριο είναι το ολικό τετραγωνικό σφάλμα:

$$J = \sum_{i=1}^{N} E(i)$$

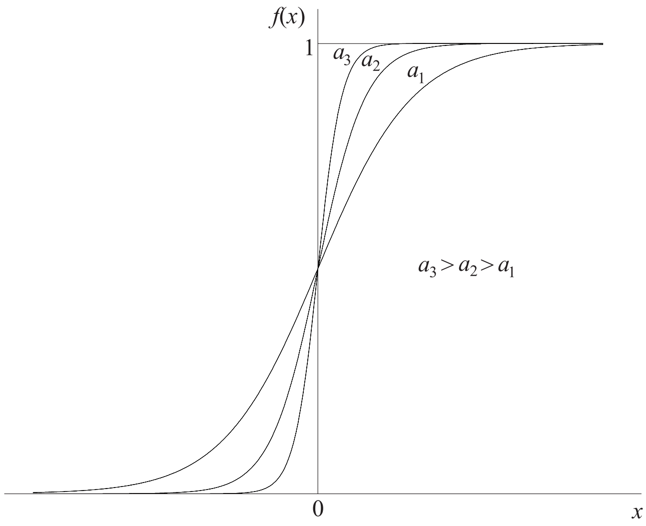
όπου

$$E(i) = \frac{1}{2}\sum_{p=1}^{m} (\hat{y}_p(i) - y_p(i))^2$$

- Η ελαχιστοποίηση του κόστους J με τη βοήθεια αναδρομικών αλγορίθμων απαιτεί την παραγώγιση του κόστους ως προς τις άγνωστες παραμέτρους, δηλαδή τις συνάψεις και τις παραμέτρους κατωφλίου όλων των νευρώνων όλων των στρωμάτων. Αυτό όμως δεν είναι δυνατόν, όταν οι συναρτήσεις ενεργοποίησης των νευρώνων είναι η ασυνεχής μοναδιαία βηματική συνάρτηση. Η δυσκολία αυτή μπορεί να ξεπεραστεί εάν στη θέση της μοναδιαίας βηματικής υιοθετήσουμε τη **λογιστική** *(logistic)* συνάρτηση που ορίζεται ως:

$$f(x) = \frac{1}{1 + \exp(-\alpha x)}$$

όπου η παράμετρος α καλείται κλίση της συνάρτησης. Όσο πιο μεγάλη η τιμή της τόσο πιο πολύ πλησιάζει η f(x) τη μοναδιαία βηματική, Σχ. 8.14. Η προσέγγιση της βηματικής από τη λογιστική συνάρτηση είναι μία από πολλές άλλες δυνατές επιλογές [2].



Σχήμα 8.14: *Η λογιστική συνάρτηση για διάφορες τιμές του α.* .

- Ο αναδρομικός αλγόριθμος ελαχιστοποίησης που παρουσιάζουμε, ανήκει στην κατηγορία των αλγορίθμων **απότομης κατάδυσης** *(gradient descent)* [2], και είναι της μορφής

$$\mathbf{w}_j^q(t+1)=\mathbf{w}_j^q(t)+\Delta\mathbf{w}_j^q$$

όπου $\mathbf{w}_j^q$ είναι το διάνυσμα των παραμέτρων (συνάψεις και κατώφλι) του j-νευρώνα στο q στρώμα και t το παρόν βήμα αναδρομής του αλγορίθμου. Η διόρθωση $\Delta\mathbf{w}_j^q$ είναι ανάλογη του gradient του κόστους J ως προς $\mathbf{w}_j^q(t)$. Λόγω της πολυπλοκότητος ενός πολυστρωματικού Perceptron ο υπολογισμός του gradient δεν είναι απλός. Ο βασικός λόγος είναι ότι οι τιμές των παραμέτρων των νευρώνων ενός στρώματος επηρεάζονται από τις τιμές των παραμέτρων των νευρώνων των προηγούμενων στρωμάτων. Ο υπολογισμός των gradients επιτυγχάνεται υπολογίζοντας πρώτα τα gradients ως προς τους νευρώνες του τελευταίου στρώματος, στη συνέχεια του προτελευταίου, κ.ο.κ. Αυτός είναι ο λόγος που ο αλγόριθμος είναι γνωστός ως ‘Αλγόριθμος Οπισθοδρομικής Διάδοσης’. Οι λεπτομέρειες είναι αρκετά τεχνικές και παραλείπονται. Ο ενδιαφερόμενες αναγνώστης παραπέμπεται στο [2].

Παρατηρήσεις

1. Η συνάρτηση κόστους J είναι μία μη γραμμική συνάρτηση ως προς τις αγνώστους παραμέτρους. Επομένως η συνάρτηση αυτή αποτελείται, γενικά, από ένα αριθμό τοπικών ελαχίστων. Ο αλγόριθμος συγκλίνει σε ένα από αυτά, ανάλογα από το σημείο εκκίνησης στο βήμα t=0. Στην πράξη, ο αλγόριθμος εφαρμόζεται παραπάνω από μία φορά, με διαφορετικές αρχικές συνθήκες και επιλέγεται η λύση που αντιστοιχεί στο μικρότερο τελικό κόστος.
2. Υπάρχει ένας αριθμός παραλλαγών του αλγορίθμου αυτού με στόχο πάντα την επιτάχυνση της σύγκλισης του [2, 3].
3. Το κόστος ελαχίστων τετραγώνων δεν είναι η μόνη επιλογή. Άλλα κόστη έχουν επίσης προταθεί και χρησιμοποιηθεί [2, 3].
4. Ένα σημαντικό πρόβλημα με τα νευρωνικά δίκτυα είναι η επιλογή του μεγέθους της αρχιτεκτονικής. Πόσα στρώματα και πόσοι νευρώνες για κάθε στρώμα απαιτούνται; Όπως για κάθε ταξινομητή, έτσι και τα δίκτυα αυτά θα πρέπει να είναι “αρκετά” μεγάλα για να μαθαίνουν το τι διαφοροποιεί τις διαφορετικές κλάσεις, αλλά όχι “τόσο μεγάλα” ώστε να μαθαίνουν το τι διαφοροποιεί τα δείγματα της ίδιας κλάσης και επίσης τις ιδιαιτερότητες του συγκεκριμένου συνόλου εκπαίδευσης. Όταν το δίκτυο είναι πολύ μεγάλο (αριθμός παραμέτρων μεγάλος) έχει περιορισμένες δυνατότητες γενίκευσης. Με άλλα λόγια, η συμπεριφορά του είναι φτωχή όταν του δίνονται διανύσματα

εκτός του συνόλου εκπαίδευσης. Για το σκοπό αυτό έχουν αναπτυχθεί μεθοδολογίες υπολογισμού ενός βέλτιστου αριθμού νευρώνων. Η πιο συνηθισμένη τάση είναι να ξεκινάμε από μεγάλα δίκτυα και να αφαιρούμε σταδιακά νευρώνες σύμφωνα με κάποιο κριτήριο [2, 3]

5. Τα πολυστρωματικά Perceptrons που περιγράψαμε δεν είναι τα μόνα νευρωνικά δίκτυα. Η βιβλιογραφία είναι πλούσια σε διαφορετικές μορφές και αρχιτεκτονικές. Εμείς, στο βιβλίο αυτό, απλώς ανοίξαμε μια πόρτα …

Άσκηση Αυτοαξιολόγησης 2/Κεφαλαίου 8

Τι είναι σωστό και τι λάθος από τα παρακάτω;

	Σωστό	Λάθος
1. Με τον αλγόριθμο Perceptron είναι δυνατή πάντα η λύση ενός προβλήματος ταξινόμησης με δύο κλάσεις.		
2. Η συνάρτηση ενεργοποίησης του Perceptron δέχεται για είσοδο το διάνυσμα των χαρακτηριστικών και δίνει έξοδο 1, ή 0.		
3. Είναι δυνατή η ταξινόμηση του XOR αν χρησιμοποιηθούν δύο ευθείες.		
4. Ένα Perceptron τριών στρωμάτων είναι δυνατόν να διαχωρίσει σε δύο κλάσεις οποιαδήποτε ένωση πολυεδρικών περιοχών στο χώρο των χαρακτηριστικών διανυσμάτων.		
5. Η Εκπαίδευση ενός πολυστρωματικού δικτύου σημαίνει τον καθορισμό της τιμής των συνάψεων και κατωφλίου για κάθε Perceptron που περιλαμβάνεται σε αυτό.		

Σύνοψη Κεφαλαίου

Στο κεφάλαιο αυτό παρουσιάσαμε το βασικό δομικό στοιχείο των νευρωνικών δικτύων, τον νευρώνα ή Perceptron. Διαπιστώσαμε ότι με τη χρήση ενός Perceptron μπορεί να διαχωριστούν δύο κλάσεις εφόσον το πρόβλημα επιδέχεται γραμμική λύση. Στη συνέχεια παρουσιάστηκε ο τρόπος ενσωμάτωσης του Perceptron σε αρχιτεκτονικές δύο και τριών στρωμάτων για την επίλυση προβλημάτων ταξινόμησης, που απαιτούν μη γραμμικές λύσεις. Τέλος παρουσιάσαμε την φιλοσοφία του αλγορίθμου οπισθοδρομικής διάδοσης, μια από τις γνωστές διαδικασίες που χρησιμοποιούνται για τον καθορισμό της τιμής των παραμέτρων των Perceptrons στις αρχιτεκτονικές των στρωμάτων.

Βιβλιογραφία

[1] F. Rosenblatt: "The Perceptron: A Probabilistic Model for Information Storage and Orangization in the Brain", Psychological Review, vol. 95, pp 386-408, 1958.

[2] S. Theodoridis, K. Koutroumbas: "Pattern Recognition", Academic Press, 1998

[3] S. Haykin: "Neural Networks", McMillan, 1994

Απαντήσεις των Ασκήσεων Αυτοαξιολόγησης

8.1

Τα διανύσματα εκπαίδευσης παρουσιάζονται στον αλγόριθμο με την ακόλουθη σειρά: $(0, 0)^T$, $(0, 1)^T$, $(1, 0)^T$, $(1, 1)^T$. Τα διανύσματα επεκτείνονται, σύμφωνα με τα όσα είπαμε στη θεωρία, προσθέτοντας το 1 ως τρίτο στοιχείο, ώστε να γίνουν οι υπολογισμοί στον 3-διάστατο χώρο για επίπεδα που διέρχονται από την αρχή των αξόνων.

Τα διαδοχικά βήματα των αναδρομών είναι:

$\mathbf{w}(0)=(0, 1, 0)^T$

ΥΠΟΛΟΓΙΣΜΟΣ ΠΡΟΣΗΜΟΥ ΤΗΣ ΠΑΡΑΣΤΑΣΗΣ $\underline{\mathbf{w}^T\mathbf{x}}$	ΑΠΟΤΕΛΕΣΜΑ ΤΑΞΙΝΟΜΗΣΗΣ	Η ΤΙΜΗ ΤΟΥ w ΓΙΑ ΤΗΝ ΕΠΟΜΕΝΗ ΑΝΑΔΡΟΜΗ
$\mathbf{w}^T(0)\ \mathbf{x}_{(0)}=(0, 1, 0)\ (0, 0, 1)^T=0$	ΛΑΘΟΣ	$\mathbf{w}(1)= \mathbf{w}(0)+(0, 0, 1)^T =(0, 1, 1)^T$
$\mathbf{w}^T(1)\ \mathbf{x}_{(1)}=(0, 1, 1)\ (0, 1, 1)^T=2>0$	ΣΩΣΤΟ	$\mathbf{w}(2)= \mathbf{w}(1) =(0, 1, 1)^T$
$\mathbf{w}^T(2)\ \mathbf{x}_{(2)}=(0, 1, 1)\ (1, 0, 1)^T=1>0$	ΛΑΘΟΣ	$\mathbf{w}(3)= \mathbf{w}(2) -(1, 0, 1)^T=(-1,1,0)$
$\mathbf{w}^T(3)\ \mathbf{x}_{(3)}=(-1,1,0)\ (1, 1, 1)^T=0$	ΛΑΘΟΣ	$\mathbf{w}(4)= \mathbf{w}(3) -(1, 1, 1)^T=(-2,0,-1)$
$\mathbf{w}^T(4)\ \mathbf{x}_{(4)}=(-2,0,-1)\ (0, 0, 1)^T=0$	ΛΑΘΟΣ	$\mathbf{w}(5)= \mathbf{w}(4)+(0, 0, 1)^T =(-2,0, 0)^T$
$\mathbf{w}^T(5)\ \mathbf{x}_{(5)}=(-2,0, 0)\ (0, 1, 1)^T=0$	ΛΑΘΟΣ	$\mathbf{w}(6)= \mathbf{w}(5)+(0, 1, 1)^T =(-2,1, 1)^T$
$\mathbf{w}^T(6)\ \mathbf{x}_{(6)}=(-2,1, 1)\ (1, 0, 1)^T=-1<0$	ΣΩΣΤΟ	$\mathbf{w}(7)= \mathbf{w}(6) =(-2,1,1)$
$\mathbf{w}^T(7)\ \mathbf{x}_{(7)}=(-2,1,1\ )\ (1, 1, 1)^T=0$	ΛΑΘΟΣ	$\mathbf{w}(8)= \mathbf{w}(7) -(1, 1, 1)^T=(-3,0,0)$
$\mathbf{w}^T(8)\ \mathbf{x}_{(8)}=(\ -3,0,0)\ (0, 0, 1)^T=0$	ΛΑΘΟΣ	$\mathbf{w}(9)= \mathbf{w}(8)+(0, 0, 1)^T =(-3,0, 1)^T$
$\mathbf{w}^T(9)\ \mathbf{x}_{(9)}=(-3,0, 1)\ (0, 1, 1)^T=1>0$	ΣΩΣΤΟ	$\mathbf{w}(10)= \mathbf{w}(9) =(-3,0,1)^T$
$\mathbf{w}^T(10)\ \mathbf{x}_{(10)}=(-3,0, 1)\ (1, 0, 1)^T=-2<0$	ΣΩΣΤΟ	$\mathbf{w}(11)= \mathbf{w}(10) =(-3,0, 1)$
$\mathbf{w}^T(11)\ \mathbf{x}_{(11)}=(-3,0, 1\ )\ (1, 1, 1)^T-2<0$	ΣΩΣΤΟ	$\mathbf{w}(12)= \mathbf{w}(11) =(-3,0, 1)$
$\mathbf{w}^T(12)\ \mathbf{x}_{(12)}=(\ -3,0,1)\ (0, 0, 1)^T=1>0$	ΣΩΣΤΟ	<u>ΤΕΛΟΣ ΥΠΟΛΟΓΙΣΜΩΝ</u>

Άρα έχει επιτευχθεί σύγκλιση και η ευθεία με παραμέτρους $\mathbf{w}=(-3, 0, 1)^T$, δηλαδή η ευθεία $-3x_1+1=0$ ταξινομεί τα διανύσματα.

Όποιος εκτέλεσε σωστά τα βήματα του αλγορίθμου αξίζει συγχαρητήρια! Αν δεν τα καταφέρατε, αλλά νομίζετε ότι θυμάστε πλήρως τον αλγόριθμο, προσπαθήστε ξανά με περισσότερη προσοχή και χωρίς να βλέπετε τη λύση. Αν δεν συμβαίνει αυτό το τελευταίο μη στενοχωριέστε, διαβάσετε το παράδειγμα 8.1 και τη σχέση (8.6) και δοκιμάστε ξανά.

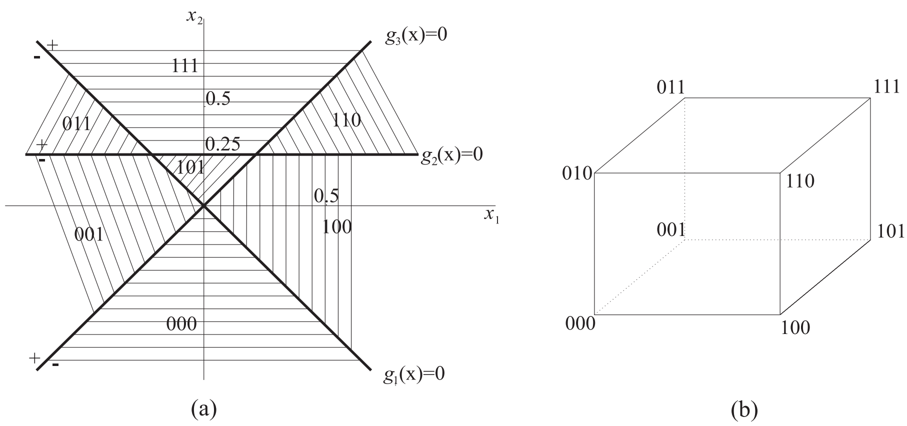
8.2

Οι τρεις ευθείες φαίνονται στο Σχ. 8.15a. Οι περιοχές που δημιουργούν έχουν ονοματιστεί σύμφωνα με τη θέση κάθε περιοχής ως προς καθεμιά από τις ευθείες. Στο Σχ. 8.15b φαίνονται οι κορυφές του κύβου στις οποίες μετασχηματίζονται οι περιοχές.

- Για να είναι το πρόβλημα διαχωρίσιμο από ένα Perceptron 2-στρωμάτων, θα πρέπει οι περιοχές να συνδυαστούν έτσι ώστε οι αντίστοιχες κορυφές του κύβου να είναι γραμμικά διαχωρίσιμες. Ένας τέτοιος συνδυασμός, προφανώς όχι ο μόνος, είναι:

$$\omega_1 : (000) \cup (001)$$

ω_2: Οι υπόλοιπες

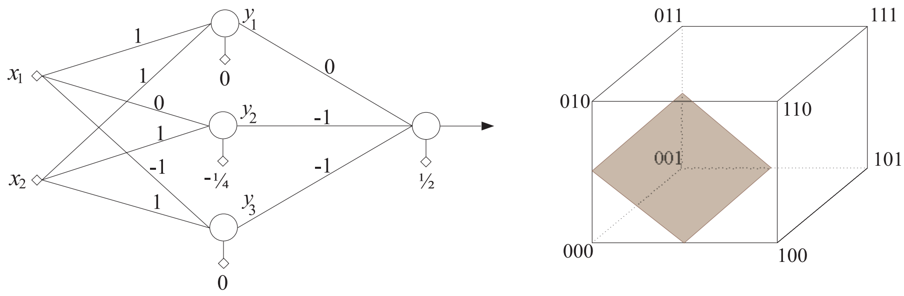


Σχήμα 8.15: *Γεωμετρία της ασκήσεως 8.2.*

Ένα επίπεδο που χωρίζει τις κορυφές αυτές είναι το

$$y_2 + y_3 - \frac{1}{2} = 0$$

που φαίνεται στο σχήμα 8.16b. Το Perceptron του σχήματος 8.16a, υλοποιεί τις ευθείες του 8.15a και το παραπάνω επίπεδο και, επομένως, επιλύει το πρόβλημα.



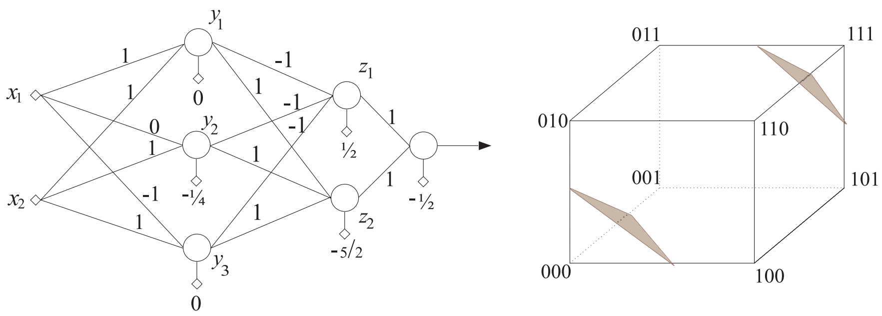
(a) (b)

Σχήμα 8.16: *Perceptron 2 στρωμάτων για την άσκηση Αυτοαξιολόγησης 8.2.*

- Ένας συνδυασμός που οδηγεί σε μη γραμμικά διαχωρίσιμες περιοχές είναι ο

$$\omega_1 : (000) \cup (111)$$

$$\omega_2\text{: Οι υπόλοιπες}$$



(a) (b)

Σχήμα 8.17: *Perceptron 3 στρωμάτων για την άσκηση Αυτοαξιολόγησης 8.2.*

Στο σχήμα 8.17b φαίνονται τα δύο επίπεδα $-y_1-y_2-y_3+1/2=0$ και $y_1+y_2+y_3-5/2=0$.
που διαχωρίζουν τις κορυφές αυτές από τις υπόλοιπες, Πράγματι, το πρώτο από τα επίπεδα αφήνει την κορυφή (000) στην (+) πλευρά και όλες τις άλλες κορυφές στην (-) πλευρά. Το δεύτερο επίπεδο αφήνει στην (+) πλευρά την κορυφή (111) και όλες τις άλλες στην (-). Ο νευρώνας εξόδου πραγματοποιεί μία OR πύλη και υλοποιεί την ευθεία $z_1 + z_2 - 1/2 = 0$. Το Perceptron του σχήματος 8.17a επιλύει το πρόβλημα.

Αξίζει συγχαρητήρια σε όποιον έλυσε ολόκληρη την άσκηση. Επιτυχία είναι και η λύση του πρώτου μέρους της. Αν δεν καταφέρατε να λύσετε ούτε το πρώτο μέρος μην στενοχωριέστε! Διαβάστε προσεκτικά την υποενότητα 8.2.3 και προσπαθήστε ξανά.

8.3

Σωστό	Λάθος

1. Με τον αλγόριθμο Perceptron είναι δυνατή πάντα η λύση ενός προβλήματος ταξινόμησης με δύο κλάσεις. Η λύση είναι δυνατή μόνο όταν οι κλάσεις είναι γραμμικά διαχωρίσιμες. (Υποενότητα 8.1)		✓
2. Η συνάρτηση ενεργοποίησης του Perceptron δέχεται για είσοδο το διάνυσμα των χαρακτηριστικών και δίνει έξοδο 1, ή 0. Οι τιμές των ουσιωδών χαρακτηριστικών πολλαπλασιάζονται επί τις συνάψεις, τα επιμέρους γινόμενα αθροίζονται με το κατώφλι και το αποτέλεσμα της άθροισης οδηγείται ως είσοδος στην f(.). (Υποενότ. 8.1.2)		✓
3. Είναι δυνατή η ταξινόμηση του XOR αν χρησιμοποιηθούν δύο ευθείες. (Ενότητα. 8.2)	✓	
4. Ένα Perceptron τριών στρωμάτων είναι δυνατόν να διαχωρίσει σε δύο κλάσεις οποιαδήποτε ένωση πολυεδρικών περιοχών στο χώρο των χαρακτηριστικών διανυσμάτων. Πράγματι! Με το πρώτο στρώμα ο χώρος μετασχηματίζεται στις κορυφές υπερκύβου, το δεύτερο στρώμα περιέχει νευρώνες για την μία μόνο κλάση και το τρίτο αποτελείται από ένα νευρώνα που υλοποιεί μια OR πύλη. (Υποενότ. 8.2.3)	✓	
5. Η Εκπαίδευση ενός πολυστρωματικού δικτύου σημαίνει τον καθορισμό της τιμής των συνάψεων και κατωφλίου για κάθε Perceptron που περιλαμβάνεται σε αυτό. Πράγματι! Ο καθορισμός αυτός γίνεται με τη χρήση γνωστών εκ των προτέρων προτύπων για κάθε κλάση και την εκτέλεση κατάλληλου αλγορίθμου ταξινόμησης. (Υποενότ. 8.2.4)	✓	

Συγχαρητήρια στον αναγνώστη που έχει απαντήσει σωστά και στις πέντε ερωτήσεις. Καλή επίδοση είναι και οι τέσσερις σωστές απαντήσεις. Όσοι έκαναν περισσότερα από δύο λάθη να μήν απογοητευτούν! Να προσπαθήσουν ξανά διαβάζοντας τις αντίστοιχες ενότητες. Είναι σίγουρο ότι η επανάληψη αυτή θα τους επιτρέψει να κατανοήσουν πλήρως και το κεφάλαιο αυτό.